

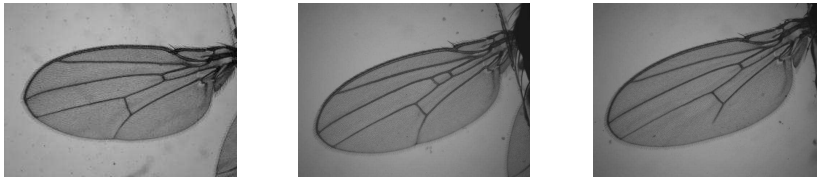
# FRUIT FLIES AND MODULI: INTERACTIONS BETWEEN BIOLOGY AND MATHEMATICS

EZRA MILLER

Possibilities for using geometry and topology to analyze statistical problems in biology raise a host of novel questions in geometry, probability, algebra, and combinatorics that demonstrate the power of biology to influence the future of pure mathematics. This is a tour through some biological explorations and their mathematical ramifications.

**A biological hypothesis.** Evolution sometimes results in discrete morphological differences among populations that diverge from a common source. This “saltation” can occur with features quantified by integers—limbs, segments, petals, teeth, or digits (humans are occasionally born with six fingers); or quantified by other discrete invariants, such as tessellation patterns—seeds in flowers or protomers in virus capsids. Biology has explanations of how populations that already exhibit a varying trait can lead to populations in which one or the other dominates. The question here is: what mechanism generates topological variation in sufficient quantity for selection to act?

Take the fruit fly, for example. The normal *Drosophila melanogaster* wing depicted on the left differs from the abnormal other two in topology as well as geometry.



Indeed, mathematically the veins in each wing can be abstracted as an embedded planar graph, with a location for each vertex and a contour for each arc. The graph in the middle has an extra edge, and hence two extra vertices, while the graph on the right is lacking a vertex. These topological variants, along with many others, occur in natural *D. melanogaster* populations, but rarely. On the other hand, different species of *Drosophila* exhibit a range of wing vein topologies. How did that come to be? Wing veins serve several key purposes, as structural supports as well as conduits for airways, nerves, and blood cells, among other things [8]. Is it possible that some force causes aberrant vein topologies to occur more frequently than would otherwise be expected in a natural population—frequently enough for evolutionary processes to act?

Results from biologist Kenneth Weber, and later with more power by David Houle’s lab, show that selecting for continuous wing deformations results in skews toward

---

*Date:* 31 July 2015.

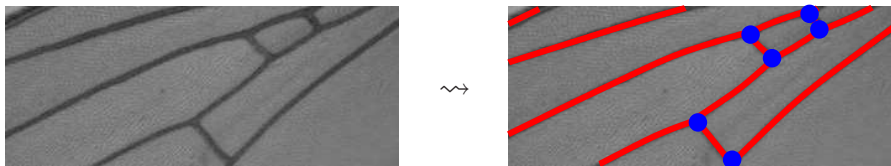
*2010 Mathematics Subject Classification.* Primary: ??X##, ??X## Secondary: ??X##, ??X##.

*Key words and phrases.* ..., ...

deformed wings with normal vein topology [39, 40, 20], but also—unexpectedly—much higher rates of topological novelty. This latter claim, which is unpublished and has yet to be tested statistically, suggests a fundamental biological hypothesis: topological novelty arises at the extreme of selection for continuous shape characteristics.

**Wings to modules.** This hypothesis could potentially be tested using persistent homology, a tool for data analysis that uses computational topology to assign modules over polynomial rings to subsets  $X \subseteq \mathbb{R}^n$  [12]. This tool is a good candidate because of its ability to emphasize differences in stratification among otherwise similar subsets of  $\mathbb{R}^n$ .

Take our case an of embedded graph  $X \subseteq \mathbb{R}^2$ , for example. For any nonnegative real numbers  $r$  and  $s$ , let  $X_r^s \subseteq X$  be the set of points at distance at least  $r$  from every vertex and within  $s$  of some edge. Thus  $X_r^s$  is obtained by taking the union of the balls of radius  $r$  around the vertices away from the union of  $s$ -neighborhoods of the edges. In the following magnified portion of the middle wing,  $r$  is approximately twice  $s$ :



The homology  $H_i(X_r^s)$  with coefficients in a field  $\mathbb{k}$  counts connected components or loops of  $X_r^s$ , when  $i = 0$  or  $1$ , respectively. Introducing a new vertex to an edge of  $X$  tends to create connected components and destroy cycles when  $r \gg s$ , because the balls around vertices protect them from the expanding edges. The precise relations between  $r$  and  $s$  that alter the topology of  $X_r^s$  depend on the geometry of  $X$ , such as the angles between edges at a given vertex, and that is the point: persistent homology uses topological invariants as measures of geometry.

To keep the data structure finite, the parameters  $r$  and  $s$  can, without loss of significant information, be restricted to integer multiples of a small positive length  $\varepsilon$ . The persistent homology of the graph  $X$  with the two parameters  $r$  and  $s$  is then defined to be the direct sum  $M_i(X) = \bigoplus_{r,s} H_i(X_r^s)$  of all of the homology groups. It is a bigraded module over the polynomial ring  $\mathbb{k}[x, y]$ : the variables  $x$  and  $y$  act on  $M_i(X)$  by comparing the homology of  $X_r^s$  with that of  $X_{r-\varepsilon}^s$  and  $X_r^{s+\varepsilon}$ , respectively.

Persistent homology with only one parameter [37, 15], instead of two or more, results in a module over a polynomial ring in one variable. This case is much better studied, in part because it behaves more tamely. In particular, there is a finite, computable set of topological features—connected components, loops or, in the general case, features of higher dimension—each of which has well defined parameters where its “birth” and “death” occur, such that every homology class is a direct sum of these features [14]. For fly wings, or arbitrary multiparameter situations, where the homology groups record the topology of several increasing chains of subsets of a single topological space, no such

clean description is possible [12]. However, alternative presentations of modules from combinatorial commutative algebra [31] (or see [33, Chapter 11]), based essentially on the theory of primary decomposition, can be understood topologically in terms of birth and death parameters. Such understanding is necessary if statistics on sets of bigraded modules are to be interpretable biologically.

**Moduli as statistical sample spaces.** Persistent homology summarizes a sample of fly wings by transforming it into a sample of modules. It is a general principle of statistics that to analyze samples from a set of objects one needs sufficient understanding of the set of all objects from metric, probabilistic, and sometimes combinatorial perspectives. How far apart are pairs of objects? How likely is each object to be selected at random? Mathematics excels at placing coherent structures on sets of all objects of a given type. The resulting “moduli spaces” pervade geometry of many sorts—differential, algebraic, arithmetic, complex, discrete—and also theoretical physics and topology, though in the latter field they are called classifying spaces. But despite their ubiquity and in some cases our substantial understanding of their geometry and combinatorics, less is known about the probability and statistics of sampling from them.

Like many moduli spaces, the ones parametrizing bigraded modules over  $\mathbb{k}[x, y]$  are quotients of algebraic varieties by continuous group actions [12]. This makes the moduli spaces complicated unions of manifolds of varying dimension. One possibility, covered in the next section, is to develop geometric methods to analyze samples from such “stratified spaces”. Another, which tends to be favored a priori for computational reasons, is to use discrete invariants as proxies for the continuous moduli. For bigraded modules, these discrete invariants include

- single-parameter persistence by tracing zigzags through the groups  $H_i(X_r^s)$  [11];
- Hilbert series, meaning the dimensions of the vector spaces  $H_i(X_r^s)$ , disregarding all of the homomorphisms between them;
- rank invariants, which take into account the ranks of the homomorphisms but not their precise algebraic structure [12, 10];
- Betti numbers, which record discrete homological invariants of the module [29].

Any discrete invariant subdivides the moduli space into regions where the invariant is constant. Understanding the nature of these subdivisions is a pragmatic matter for bigraded  $\mathbb{k}[x, y]$ -modules, given the fly wing context, but (modifications of) some of these new sorts of questions make sense for any moduli space, or indeed any stratified space.

1. What metric or combinatorial properties do these subdivisions possess? For example, do the regions have equal dimension and roughly equal size, or are there a few big regions (or only one) of top dimension and bunch of smaller ones?
2. What distribution of discrete invariants are expected from a given (biological) problem? Might the discrete invariants be expected to distinguish between the modules produced by applied situations even if the regions aren’t of similar size?

3. General geometric statistical question: what (natural) measures should be placed on a set of discrete invariants, given the geometry of the moduli spaces?
4. Can the continuous variation be captured discretely to desired precision? More precisely, is there a family (indexed by  $n$ ) of sets of discrete invariants such that letting  $n \rightarrow \infty$  results in an increasingly fine subdivision?

**Geometric probability on stratified spaces.** As we have seen for fly wings, statistical problems where the sample objects are more complicated than vectors in vector spaces naturally lead to sampling from stratified spaces. The goal of geometric statistics in this setting is, like in ordinary linear statistics, to identify, describe, summarize, or make inferences about an unknown probability distribution on the sample space from which the sample points are assumed to be drawn. To that end, it is crucial to understand the opposite problem, from probability theory: given a distribution on the relevant sample space, how do samples from that distribution behave?

The simplest summary of a distribution is a point—an average or *population mean*—about which the distribution is centered. Laws of large numbers assert that means of increasingly large random samples from a distribution converge to a population mean. Statistics requires knowledge of the expected difference between a sample mean and population mean. Central limit theorems help quantify that difference by describing the variation of sample means around population means. In ordinary statistics, for example, when the sample space is the real line, the central limit theorem dictates that sample means vary around the population mean according to a distribution that is, in the limit of infinite sample size, Gaussian.

In Euclidean space, basic concepts such as mean, expectation, and average coincide and therefore admit multiple equivalent characterizations, such as via least squares or arithmetic average. Already thinking about asymptotics of samples from smooth manifolds—let alone singular spaces such as the moduli spaces relevant to fly wings—requires a radical shift in perspective, as compared with samples from linear spaces, because different characterizations lift to different notions in curved settings [23]. Moreover, for many of these notions, such as Fréchet mean defined by least squares, the minimizer is not unique: what is the average of the north pole and south pole on the sphere? It is the entire equator. (This explains the phrase “a population mean” in the previous paragraph, as opposed to “the population mean”.) Nonetheless, laws of large numbers hold [41, 5], and central limit theorems exist in various situations [26, 17, 18, 6, 22], such as when the data are concentrated near a Fréchet mean. (Many of these theorems were motivated by biology; read the title of [22], for instance.) For additional background and references concerning statistics on manifolds, see [23].

Statistics on smooth manifolds relies on approximation of the manifold by its tangent space, which is Euclidean. Once a metric on the manifold has been specified—a necessary and often nontrivial step for statistics, because of the need to know how far apart sample objects are—the exponential map at a point  $x$  takes a neighborhood  $U$

of 0 in the tangent space  $T_x$  to a neighborhood  $\exp(U)$  around  $x$ . Ordinary probability in the vector space  $T_x$  is transformed into geometric probability on the smooth sample space via the exponential map at  $x$ , which is close to an isometry when  $U$  is small. In particular, central limit theorems on smooth manifolds can be interpreted as describing variation of sample means around a population mean by pushing forward the linear setup along an exponential map.

In the singular setting of stratified spaces, there is no general method to compare with or reduce to ordinary linear probability and statistics in the tangent space at a mean. The types of sample spaces  $M$  intended here are those possessing a *topological stratification* (see [16] or [35]): an expression as a disjoint union  $M = M_0 \cup M_1 \cup \dots \cup M_r$  of *strata* such that for all  $d \in \{0, \dots, r\}$ ,

- the stratum  $M_d$  is a manifold,
- $M_0 \cup \dots \cup M_d$  is closed in  $M$ , and
- for every pair  $x, y \in M_d$  there is a homeomorphism  $M \rightarrow M$  that takes  $x$  to  $y$  and takes each stratum to itself.

The third condition ensures that the topology of  $M$  and its stratification behaves precisely the same way near  $x$  as it does near  $y$ . Examples include graphs (or networks), whose strata are vertices and edges; polytopes, whose strata are (relatively open) faces; and real (semi)algebraic varieties, whose strata consist of classes of singular points. The tangent space  $T_x$  to a stratified space  $M$  at a point  $x$  is a cone over a stratified space of dimension one less than  $\dim(M)$ . If  $M$  is already a cone with apex  $x$ , then  $T_x \cong M$  is as complicated as  $M$  itself; capturing the local structure of a stratified space near a point need not simplify the geometry the way it does in the smooth setting.

Cases where stratified laws of large numbers and central limit theorems are known occur in specific simple examples where comparison with linear spaces is possible, such as open books [19] (unions of Euclidean half-spaces glued along their boundary subspaces), isolated planar hyperbolic singularities (cones where the singular point has angle sum  $> 2\pi$  instead of the ice-cream case of  $< 2\pi$ ) [24], and binary trees [4]. But in general, de novo geometric constructions are required. Such has been the avenue for the good deal of probability theory, including laws of large numbers, that has been established in the generality of nonpositively curved spaces (see [38]), which are defined as spaces whose triangles with given edge lengths are thinner than would be expected from Euclidean geometry (see [9]). The metric structures of nonpositively curved spaces induce a number of simplifying consequences, such as uniqueness of Fréchet means, which have played important roles in the progress thus far in geometric probability and statistics on stratified spaces. Nonetheless, the promising interactions of nonpositive curvature with geometrically stratified probability and statistics remain in their infancy.

**Sticky means.** The novelty of attempting statistics on stratified sample spaces is exemplified by nonclassical “sticky” phenomena that can occur at singularities. In Euclidean statistics, the mean of a finite set of points moves slightly in any desired

direction by perturbing the points. This intuition extends to manifolds, by linear approximation [26, 17, 6, 22], but it can fail even in the simplest singular sample spaces. Consider the tripod, for instance, depicted at left:



In the center-left figure, the Fréchet mean  $\mu$  of the three points on the legs is the origin 0, by symmetry. But wiggling the three points, as in the center-right figure, does not move the Fréchet mean at all; one of the points would have to be moved more than twice as far from the center, as in the final figure, to nudge the mean onto its leg.

An open book with three pages is a product of tripod with vector space  $\mathbb{R}^d$ . (To get an arbitrary number of pages, replace the tripod with a graph having an arbitrary number of rays emanating from the center point. It bears mentioning that every topologically stratified space  $M$  is locally homeomorphic to an open book near any point on a stratum of dimension  $\dim(M) - 1$ . In other words, the tangent cones to points on codimension 1 strata are open books. Hence this example is universal in some sense.) In an open book, the mean sticks to the spine—the copy of  $\mathbb{R}^d$  that is contained in all three pages—when three similarly situated points are wiggled, although that wiggling can move the mean in arbitrary directions inside of the spine [19].

With these examples and others in mind, a formal definition has been developed [24]: let  $\mathcal{M}$  be a set of measures on a metric space  $K$ . Assume  $\mathcal{M}$  has a given topology. A *mean* is a continuous assignment  $\mathcal{M} \rightarrow \{\text{closed subsets of } K\}$ . A measure  $\mu$  *sticks* to a closed subset  $C \subseteq K$  if every neighborhood of  $\mu$  in  $\mathcal{M}$  contains a nonempty open subset consisting of measures whose mean sets are contained in  $C$ .

Stickiness implies that it is possible for the means of large samples from a distribution on a stratified space to lie in a subset of low dimension, with positive probability (equal to 1 in some cases, such as the tripod), even if the distribution being sampled is well behaved [2, 4, 19, 24]. This contrasts with usual laws of large numbers, where sample means approach the population mean but almost surely never land on it—or on any given subset of low dimension containing it. Thinking in terms of central limit theorems, whereas in usual cases the limiting distributions have full support, in sticky cases the limiting distributions can have components supported on low-dimensional subsets of the sample space.

Examples aside, central limit theorems on stratified spaces of any generality have yet to be formulated, let alone proved. Subtle and deep behavior associated with the boundary between sticky and non-sticky are still being discovered. In particular, the distinction between positive and negative curvature seems to be critical for stickiness. One common type of positive curvature, particularly in statistical problems, appears when a flat or positively curved smooth manifold is quotiented by a proper Lie group



action. Shape spaces (see [27, 28])—including those applicable to fly wings with constant topology by keeping track of the locations of the vertices of the graph—have this form, for instance, being quotients of matrix spaces by actions of rotations, scaling, or projective transformations. Huckemann [23] has shown essentially that when sampling from a stratified space that is a quotient of this form, Fréchet means run away from singularities and hence land almost surely in the smooth locus. On the other hand, every case exhibiting stickiness has negative curvature (in the sense of Alexandrov: curvature bounded strictly above by 0; see [9]). It remains an open problem to formulate a condition, in terms of something like negative sectional curvature, that allows means to run toward singularities of the space for appropriate types of distributions on it.

Further potential implications for pure mathematics emerge when considering how to recover curvature invariants from asymptotics instead of the other way around. In the smooth case, for local samples—that is, sufficiently nearby the mean—accounting for curvature reduces central limit theorems to the Euclidean version. Thus, if the curvature is unknown but properties of the distribution are known, then the curvature ought to be recoverable. In singular settings, attempting such a recovery could give rise to singular analogues of smooth Riemannian curvature invariants.

**Phylogenetic trees.** Moduli spaces of fly wing modules constitute just one of myriad ways that geometric probability on stratified spaces can arise. Among those, nothing is special about biology, except perhaps that its diversity of forms and the nature of their variation lend themselves to geometric data analysis of this sort. That said, the genesis of stratified statistics came directly from another evolutionary biology moduli space.

One of the principal aims of systematic and evolutionary biology is to determine relationships between species. Trees representing these kinships are reconstructed from biological data such as DNA sequences or morphology. Experimental procedures generate distributions on the set of phylogenetic trees in multiple ways. For example, the evolutionary history of a single gene across multiple individuals is represented by a “gene tree”. Natural processes such as incomplete lineage sorting cause gene trees sampled from a set of individuals to differ in topology from one another and from the evolutionary history of their set of species—the history of population bifurcations leading to divergence (see [30], for example). Another crucial example occurs not from the data but from the method of inference: the problem of phylogenetic tree reconstruction is intractable enough that probabilistic methods are commonly used, resulting in posterior distributions instead of a single optimum [36, 25].

Mathematically speaking, a phylogenetic tree on a given set of species is a connected metric graph, with no loops, whose vertices of degree 1 (“leaves”) are labeled by the species. The introduction by Billera, Holmes, and Vogtmann of an appropriate moduli space for the problem, namely the space of phylogenetic trees [7], initiated a surge of activity attempting to mine the combinatorics and geometry of the space to devise statistical methods. And the combinatorics is formidable: for  $n$  species, the tree space

$\mathcal{T}_n$  is a polyhedral stratified space composed of  $(2n - 3)!!$  Euclidean orthants of dimension  $n - 2$ . Despite its complexity,  $\mathcal{T}_n$  succumbed: the advent of a polynomial time algorithm for shortest paths in tree space [34] made it possible to compute Fréchet means in  $\mathcal{T}_n$  [1, 32] based on probability theory for nonpositively curved spaces [38].

The geometric probability on stratified spaces in the previous two sections was initially developed specifically to understand the behavior of Fréchet means in the moduli space  $\mathcal{T}_n$ . The simple examples [19, 24] deal with informative local subsets of  $\mathcal{T}_n$ . In addition to those, efforts are underway to prove laws of large numbers and central limit theorems in the global context of  $\mathcal{T}_n$  itself [2, 3].

Stickiness in tree space has a concrete, meaningful biological interpretation (although the jury is out on the extent to which the interpretation reflects reality). Points in strata of lower dimension represent phylogenetic trees with one or more non-binary vertices. Biologically, these are unresolved phylogenies: one species diverges simultaneously into three new species, for example, instead of first splitting into two new species followed by another binary divergence event from one of the two. Sets of phylogenetic trees from biological experiments often contain evidence for many or all of the possible sequences of binary divergence events that resolve a given multiple divergence. Stickiness implies that the mean tree will contain an unresolved divergence whenever there is insufficient strength of pull toward any one of the resolving binary sequences to support the conclusion it represents. The picture to keep in mind is the tripod, whose stickiness we saw earlier: it is the tree space  $\mathcal{T}_3$  on three species.

**Conclusion.** Spaces of biological forms provide an environment in which mathematical methods can assign distances between phenotypes. The lines of inquiry here fit into biologist David Houle’s vision of phenomics [21], particularly the genotype–phenotype map. To make a long story short, selection acts on phenotype whereas descent and modification act via genotype, so it becomes desirable to compare phenotypic distance to genotypic distance, including some working definition of each distance in any given case study. In general, to grapple with statistical correlations between genotypes and phenotypes requires a mathematical parametrization for each of those biological concepts. For genotypes, that is likely to involve combinatorial considerations, since the basic quantum of information is discrete. Perhaps large-scale continuous analogues or approximations might be meaningful, as they are for statistical mechanics, and that is another potential departure point for mathematics. On the other hand, phenotype is often continuous in nature: what are the locations of the vertices in a fly wing? How do the arcs between the vertices bow outward or inward? How do these characteristics change from wing to wing? Parametrization in such a context requires thinking about spaces of continuous objects, the sort of thinking that mathematics is designed to carry out. The examples presented here demonstrate a sample of the kinds of abstract structures in pure mathematics—along with unexpected questions about them—that biological investigations reveal.



## REFERENCES

- [1] Miroslav Bačák, *Computing medians and means in Hadamard space*, SIAM J. Optimiz. **24** (2014), 1542–1566.
- [2] Dennis Barden, Huiling Le, and Megan Owen, *Central limit theorems for Fréchet means in the space of phylogenetic trees*, Electron. J. Probability **8** (2013), no. 25, 1–25.
- [3] Dennis Barden, Huiling Le, and Megan Owen, *Limiting behaviour of Fréchet means in the space of phylogenetic trees*, preprint, 2014. arXiv:math.PR/1409.7602v1
- [4] B. Basrak, *Limit theorems for the inductive mean on metric trees* J. Appl. Prob. **47** (2010), 1136–1149.
- [5] R. Bhattacharya and V. Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds I*, Ann. Statist. **31** (2003), no. 1, 1–29.
- [6] R. Bhattacharya and V. Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds II*, Ann. Statist. **33** (2005), no. 3, 1225–1259.
- [7] L. Billera, S. Holmes, and K. Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. in Appl. Math. **27** (2001), 733–767.
- [8] Seth S. Blair, *Wing vein patterning in Drosophila and the analysis of intercellular signaling*, The Annual Review of Cell and Developmental Biology **23** (2007), 293–319.
- [9] Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **319**, Springer-Verlag, Berlin, 1999.
- [10] Francesca Cagliari, Barbara Di Fabio, Massimo Ferri, *One-dimensional reduction of multidimensional persistent homology*, Proceedings of the American Mathematical Society **138** (2010), no. 8, 3003–3017.
- [11] Gunnar Carlsson and Vin de Silva, *Zigzag persistence*, Foundations of Computational Mathematics **10** (2010), no. 4, 367–405.
- [12] Gunnar Carlsson and Afra Zomorodian, *The theory of multidimensional persistence*, Discrete and Computational Geometry **42** (2009), 71–93.
- [13] James Degnan and Noah Rosenberg, *Discordance of species trees with their most likely gene trees*, PLoS Genetics **3** (2006), 762–768.
- [14] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian, *Topological persistence and simplification*, Discrete and Computational Geometry **28**, 511–533.
- [15] Patrizio Frosini and Claudia Landi, *Size functions and formal series*, Applicable Algebra in Engineering, Communication and Computing **12** (2001), no. 4, 327–349.
- [16] Mark Goresky and Robert MacPherson, *Stratified Morse theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], **14**, Springer-Verlag, 1988.
- [17] H. Hendriks and Z. Landsman, *Asymptotic behaviour of sample mean location for manifolds*, Statistics and Probability Letters **26** (1996), 169–178.
- [18] H. Hendriks and Z. Landsman, *Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions*, Journal of Multivariate Analysis **67** (1998), 227–243.
- [19] Thomas Hotz, Stephan Huckemann, Huiling Le, J. S. Marron, Jonathan C. Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru, and Sean Skwerer) *Sticky central limit theorems on open books*, Annals of Applied Probability **23** (2013), no. 6, 2238–2258. doi:10.1214/12-AAP899 arXiv:math.PR/1202.4267
- [20] David Houle, Jason Mezey, Paul Galpern1, and Ashley Carter, *Automated measurement of Drosophila wings*, BMC Evolutionary Biology **3** (2003), 25–37.
- [21] David Houle, *Numbering the hairs on our heads: The shared challenge and promise of phenomics*, Proceedings of the National Academy of Sciences USA **107** (2010), 1793–1799.

- [22] Stephan Huckemann, *Inference on 3D Procrustes means: Tree boles growth, rank-deficient diffusion tensors and perturbation models*, Scandinavian Journal of Statistics **38** (2011), no. 3, 424–446.
- [23] Stephan Huckemann, *On the meaning of mean shape: manifold stability, locus and the two sample test*, Annals of the Institute of Statistical Mathematics, **64** (2012), 1227–1259.
- [24] Stephan Huckemann, Jonathan C. Mattingly, Ezra Miller, and James Nolen, *Sticky central limit theorems at isolated hyperbolic planar singularities*, submitted, 35 pages. arXiv:math.PR/1410.6879
- [25] John P. Huelsenbeck and Fredrik Ronquist, *MrBayes: Bayesian inference of phylogenetic trees*, Bioinformatics **17** (2001), 754–755.
- [26] P. Jupp, *Residuals for directional data*, J. Appl. Statist. **15** (1988), no. 2, 137–147.
- [27] David G. Kendall, *Shape manifolds, Procrustean metrics, and complex projective spaces*, Bull. London Math. Soc. **16** (1984), 81–121.
- [28] David G. Kendall, Dennis Barden, Thomas K. Carne, and Huiling Le, *Shape and shape theory*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1999.
- [29] Kevin Knudson, *A refinement of multi-dimensional persistence*, Homology, Homotopy and Applications **10** (2008), no. 1, 259–281.
- [30] Wayne P. Maddison, *Gene trees in species trees*, Systematic Biology **46** (1997), 523–536.
- [31] Ezra Miller, *Cohen–Macaulay quotients of normal semigroup rings via irreducible resolutions*, Math. Res. Lett. **9** (2002), no. 1, 117–128.
- [32] Ezra Miller, Megan Owen, and Scott Provan *Polyhedral computational geometry for averaging metric phylogenetic trees*, Advances in Applied Math. **15** (2015), 51–91.
- [33] Ezra Miller and Bernd Sturmfels, *Combinatorial commutative algebra*, Graduate Texts in Mathematics, **227**, Springer-Verlag, 2005.
- [34] Megan Owen and J.S. Provan, *A fast algorithm for computing geodesic distance in tree space*, IEEE/ACM Trans. Comput. Biol. Bioinform. **8** (2011), 2–13.
- [35] Markus J. Pflaum, *Analytic and Geometric Study of Stratified Spaces*, Lecture Notes in Mathematics, **1768**, Springer-Verlag, 2001.
- [36] Bruce Rannala and Ziheng Yang, *Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method*, Molecular Biology and Evolution **14** (1997), 717–724.
- [37] Vanessa Robins, *Towards computing homology from finite approximations*, Topology Proceedings **24** (1999), no. 1, 503–532.
- [38] K.-T. Sturm, *Probability measures on metric spaces of nonpositive curvature*, in Heat kernels and analysis on manifolds, graphs, and metric spaces: lecture notes from a quarter program on heat kernels, random walks, and analysis on manifolds and graphs, Contemporary Mathematics **338** (2003), 357–390.
- [39] K. E. Weber, *Selection on wing allometry in Drosophila melanogaster*, Genetics **126** (1990), 975–989.
- [40] K. E. Weber, *How small are the smallest selectable domains of form?* Genetics **130** (1992), 345–353.
- [41] H. Ziezold, *On expected figures and a strong law of large numbers for random elements in quasi-metric spaces*, in Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A, pp. 591–602, Reidel, Dordrecht, 1977.

MATHEMATICS DEPARTMENT, DUKE UNIVERSITY, DURHAM, NC 27708

Website: <http://www.math.duke.edu/~ezra>